

Transformer-based Reasoning for Learning Evolutionary Chain of Events on Temporal Knowledge Graph

Zhiyu Fang*
Shuai-Long Lei*
mr.fangzy@foxmail.com
shuailong0lei@gmail.com
University of Science and Technology
Beijing
Haidian, Beijing, China

Xiaobin Zhu†
University of Science and Technology
Beijing
Haidian, Beijing, China
zhuxiaobin@ustb.edu.cn

Chun Yang
University of Science and Technology
Beijing
Haidian, Beijing, China
chunyang@ustb.edu.cn

Shi-Xue Zhang
University of Science and Technology
Beijing
Haidian, Beijing, China
zhangshixue111@163.com

Xu-Cheng Yin
University of Science and Technology
Beijing
Haidian, Beijing, China
xuchengyin@ustb.edu.cn

Jingyan Qin
University of Science and Technology
Beijing
Haidian, Beijing, China
qinjingyanking@foxmail.com

ABSTRACT

Temporal Knowledge Graph (TKG) reasoning often involves completing missing factual elements along the timeline. Although existing methods can learn good embeddings for each factual element in quadruples by integrating temporal information, they often fail to infer the evolution of temporal facts. This is mainly because of (1) insufficiently exploring the internal structure and semantic relationships within individual quadruples and (2) inadequately learning a unified representation of the contextual and temporal correlations among different quadruples. To overcome these limitations, we propose a novel Transformer-based reasoning model (dubbed ECEformer) for TKG to learn the Evolutionary Chain of Events (ECE). Specifically, we unfold the neighborhood subgraph of an entity node in chronological order, forming an evolutionary chain of events as the input for our model. Subsequently, we utilize a Transformer encoder to learn the embeddings of intra-quadruples for ECE. We then craft a mixed-context reasoning module based on the multi-layer perceptron (MLP) to learn the unified representations of inter-quadruples for ECE while accomplishing temporal knowledge reasoning. In addition, to enhance the timeliness of the events, we devise an additional time prediction task to complete effective temporal information within the learned unified representation. Extensive experiments on six benchmark datasets verify the state-of-the-art performance and the effectiveness of our method.

*Both authors contributed equally to this research.

†Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '24, July 14–18, 2024, Washington, DC, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0431-4/24/07

<https://doi.org/10.1145/3626772.3657706>

CCS CONCEPTS

• **Computing methodologies** → **Temporal reasoning**.

KEYWORDS

Temporal Knowledge Graph Completion, Context Information Mining, Link Prediction, Evolutionary Chain of Event

ACM Reference Format:

Zhiyu Fang, Shuai-Long Lei, Xiaobin Zhu, Chun Yang, Shi-Xue Zhang, Xu-Cheng Yin, and Jingyan Qin. 2024. Transformer-based Reasoning for Learning Evolutionary Chain of Events on Temporal Knowledge Graph. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*, July 14–18, 2024, Washington, DC, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3626772.3657706>

1 INTRODUCTION

Knowledge graphs (KGs) store human-summarized prior knowledge in structured data formats and are widely applied in many downstream tasks, such as information retrieval [2], intelligent question answering [3], and recommendation systems [44]. Traditional knowledge graphs employ triples (s, p, o) to record each fact or event in the knowledge base, where s and o respectively represent the subject and object entities, and p denotes the logical predicate (or relation) connecting these two entities. e.g., (*Barack Obama*, *presidentOf*, *USA*). However, events in the real world often dynamically evolve over time. To extend the temporal nature of events, Temporal Knowledge Graphs (TKGs) incorporate temporal information into the traditional triples, resulting in quadruples (s, p, o, τ) . e.g., (*Barack Obama*, *presidentOf*, *USA*, *2009–2017*) and (*Donald Trump*, *presidentOf*, *USA*, *2017–2021*).

Completing missing events for specific timestamps in TKGs via Temporal Knowledge Graph Reasoning (TKGR) has recently attracted growing attention from both academic and industrial communities. For example, the missing element *Barack Obama* in the quadruple query $(?, \textit{president of, USA, 2009–2017})$ can be predicted via TKGR. Existing TKGR methods can be broadly categorized into two types according to the temporal scope of link prediction,

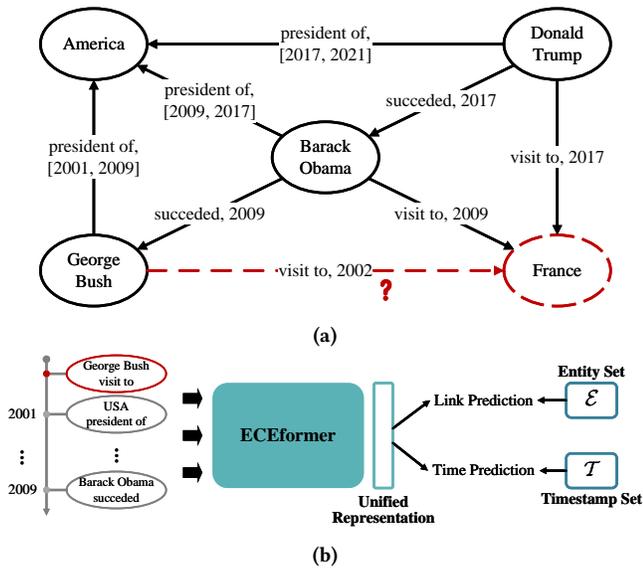


Figure 1: Illustration of the overall concept of our method. Part (a) displays a subgraph of TKG, which includes facts related to three U.S. Presidents. Part (b) illustrates our motivation for the ECEformer to learn a unified representation from an evolutionary chain of events, thereby simultaneously addressing link and time prediction tasks.

i.e., interpolation and extrapolation [16]. The former focuses on completing unknown knowledge in the past, whereas the latter concentrates on estimating unknown knowledge in the future. Concretely, interpolation reasoning methods aim to effectively integrate temporal information into the evolutionary trajectory of events. They typically utilize geographic and geometric information to model the relationships among elements within individual quadruples. Accordingly, HyTE [8] employs a translation-based strategy [4] to measure the distance between entities and relations in the customized hyperplane, where each timestamp is associated with a corresponding hyperplane. Inspired by BoxE [1], BoxTE [27] additionally incorporates temporal information embedded through the relation-specific transition matrix within the foundational box-type relation modeling. TGeomE [41], a geometric algebra-based embedding approach, performs 4th-order tensor factorization and applies linear temporal regularization for time representation learning. These methods impose a rigid prior on KGs, relying solely on geometric modeling and geographic measurements. However, they ignore the implicit semantic information, failing to fully capture the complex graph structure and relational context. Therefore, it is necessary to **sufficiently explore the internal structure and semantic relationships within individual quadruples (Limitation I)**.

On the other hand, extrapolation reasoning methods aim to effectively leverage the query-associated historical information [36]. They generally utilize Graph Neural Network (GNN) models to capture structural characteristics within each snapshot and leverage Recurrent Neural Networks (RNNs) to explore the evolutionary characteristics across timestamps. Accordingly, RE-Net

[16] employs a multi-relational graph aggregator to capture the graph structure and an RNN-based encoder to capture temporal dependencies. RE-GCN [24] leverages a relation-aware Graph Convolution Network (GCN) to capture the structural dependencies for the evolution unit and simultaneously captures the sequential patterns of all facts in parallel using gate recurrent components. RPC [25] mines the information underlying the relational correlations and periodic patterns via two correspondence units: one is based on a relational GCN, and the other is based on Gated Recurrent Units (GRUs). However, this phased processing strategy can easily degrade the accuracy of predicting future events due to incorrect historical information. Therefore, it is necessary to adaptively correct the interference of incorrect historical information on prediction accuracy by **learning a unified representation of the contextual and temporal correlations among different quadruples (Limitation II)**.

Given a subgraph of TKG (as shown in Figure 1a), interpolation methods utilize geographical information to maximize structural differences between different quadruples. However, this type of approaches cannot effectively utilize semantic information in reasoning. Moreover, learning discriminative and powerful representations for each item within a quadruple is challenging, especially in modeling timestamps. For instance, differentiating between (*Barack Obama, visitTo, France, 2009*) and (*Donald Trump, visitTo, France, 2017*) necessitates precise characterizations of different names and timestamps. Although extrapolation methods consider neighborhood information, independently processing structural and temporal characteristics can lead to confusion in reasoning. For instance, under the timestamps of *2009* and *2017*, the structure of *Barack Obama* and *Donald Trump* with *USA* and *France* is consistent. Hence, the model can naturally infer the structure of *George Bush* with *USA* and *France* in *2001* should be consistent with the structure of *Barack Obama* and *Donald Trump*, which contradicts the facts. From this key observation, we believe that designing an end-to-end network for learning a unified representation rich in information from temporal subgraph inputs will enhance the performance of TKGR. This network adaptively extracts both intra-quadruple and inter-quadruple, together with temporal information.

According to the aforementioned analysis, we propose an innovative end-to-end Transformer-based network to learn the evolutionary chain of events (dubbed ECEformer) for TKGR. Figure 1b illustrates the motivation of our method. To easily illustrate the structural and contextual relationship between the query node and the adjacent nodes, we form an evolutionary chain of events (ECE) by extracting query-specific neighbors from the subgraph. Subsequently, our proposed Transformer-based model learns the unified representation for the ECE via two novel modules, namely, ECE Representation Learning (ECER) and Mixed-Context Knowledge Reasoning (MCKR). Specifically, to overcome **Limitation I**, ECER employs a Transformer-based encoder to embed each event in the ECE, such as (*USA, president of, 2001*) corresponding to a specific query *George Bush*. To overcome **Limitation II**, MCKR induces the embeddings of each event and enhances interaction within and between quadruples via crafting an MLP-based information mixing layer. In addition, to enhance the timeliness of the events, we devise an additional time prediction task to imbue effective temporal information within the learned unified representation.

In summary, our main contributions are three-fold:

- We propose an innovative temporal knowledge graph reasoning method, namely ECEformer. Experimental results verify the state-of-the-art performance of our method on six publicly available datasets.
- Proposing a Transformer-based encoder for ECE representation learning (ECER) aims to explore the internal structure and semantic relationships within individual quadruples.
- Proposing an MLP-based layer for Mixed-Context Knowledge Reasoning (MCKR) aims to learn a unified representation of the contextual and temporal correlations among different quadruples.

2 RELATED WORK

2.1 Interpolation-based TKGR

Interpolation-based TKGR methods estimate missing facts or events by identifying consistent trends within a TKG. Based on the process strategy for temporal information, we categorize them into two types: timestamp-independent and timestamp-specific methods.

Timestamp-independent methods treat timestamps as independent units, equivalent to entities or relations, and do not apply additional operations on the timestamps. Generally, these methods [19, 20, 27, 29] directly associate timestamps to the corresponding entities and relations based on the foundation of static KG models. For example, Leblay *et al.* [20] extended the classic TransE [4] model to TKGR by concatenating timestamp embeddings with relation embeddings. Inspired by BoxE [1], Messner *et al.* [27] directly introduced timestamp embeddings on its basis. However, timestamp-independent methods are usually limited in capturing the temporal information of evolutionary events. Timestamp-specific methods embed the temporal information and learn the evolution of entities and relations via timestamp-specific functions. To effectively utilize the structural and semantic information of the time, they ingeniously designed various time-specific functions, such as diachronic embedding functions [11], time-rotating functions [5, 40], time-hyperplane functions [8, 38, 41, 47], and non-linear embedding functions [12, 17, 22, 35, 39]. Specifically, Goel *et al.* [11] proposed a diachronic entity embedding function to characterize entities at any timestamp. Chen *et al.* [5] defined the spatial rotation operation of entities around the time axis. Zhang *et al.* [47] learned spatial structures interactively between the Euclidean, hyperbolic and hyper-spherical spaces. Han *et al.* [12] explored evolving entity representations on a mixed curvature manifold using a velocity vector defined in the tangent space at each timestamp. Despite the great efforts made by these studies, they are limited in predicting future events [49].

2.2 Extrapolation-based TKGR

Extrapolation-based TKGR methods predict future facts or events by learning effective embeddings from historical snapshots. As knowledge subgraphs at specific timestamps, historical snapshots inherently contain rich structural and semantic information. Extrapolation methods typically employ deep learning techniques such as Graph Neural Networks (GNNs) [24, 45, 46], Recurrent Neural Networks (RNNs) [16, 21], and Reinforcement Learning (RL) [33, 49] to extract features from historical snapshots.

To elaborate further, GNN-based methods are naturally applied to TKGR due to the inherent topological structure of historical snapshots. Schlichtkrull *et al.* [30] early utilized Graph Convolutional Networks (GCN) to model the multi-relational features of traditional KG. To incorporate the temporal characteristics in TKGs, Jin *et al.* [16] additionally introduced an RNN-based sub-network specifically for recursively encoding past facts. Li *et al.* [24] presented the recurrent evolution network based on GCN, which learns the evolutionary representations of entities and relations at each timestamp by modeling the KG sequence recurrently. To handle the large-scale inputs and non-independent and identically distributed data, Deng *et al.* [9] combined GCN and GRU for simultaneously predicting concurrent events of multiple types and inferring multiple candidate participants. RL-based methods are becoming popular to improve the interpretability of reasoning methods, especially in question-answering tasks. By mimicking the human mechanism of searching for historical information and meticulous reasoning, CluSTeR [23] implements an interpretable prediction for TKGR using a two-stage approach of clue searching and temporal reasoning. DREAM [49] introduces an adaptive reinforcement learning model based on attention mechanism, addressing the challenges in traditional RL-based TKG reasoning, specifically the lack of capturing temporal evolution and semantic dependence jointly and the over-reliance on manually designed rewards. Furthermore, some other types of methods also achieve good performance in TKGR. For example, TLogic [26] introduces an explainable framework based on temporal logical rules, leveraging temporal random walks to provide explanations while maintaining time consistency. Faced with the challenge of predicting future facts for newly emerging entities based on extremely limited observational data, MetaTKGR [37] dynamically adjusts its strategies for sampling and aggregating neighbors from recent facts and introduces a temporal adaptation regularizer to stabilize the meta-temporal reasoning process over time. In summary, extrapolation methods generally process structural and temporal characteristics independently. This approach ignores the intrinsic correlation between structural and temporal information, limiting the capability to extract a unified representation of both features.

2.3 Transformer in Knowledge Graph

With the Transformer achieving excellent performance in numerous tasks in natural language processing and computer vision, it has been introduced into knowledge graph tasks. Leveraging the powerful learning capability of Transformers for sequential data structures and contexts, numerous Transformer-based methods [6, 7, 14, 43] have emerged for traditional KGs. For example, Chen *et al.* [6] proposed two different Transformer blocks for hierarchically learning representations of entities and relations. In addition, some methods also achieve remarkable results in downstream tasks of KGs. Specifically, Hu *et al.* [13] utilized a Transformer-based framework for encoding the content of neighbors of an entity for entity typing task. Meanwhile, Xu *et al.* [42] applied it for context-aware rule mining over KG. To serve diverse KG-related tasks, Zhang *et al.* [48] proposed a uniform knowledge representation and fusion module via structure pretraining and prompt tuning.

Transformer enables the exploration of structural associations within each historical snapshot and captures the temporal relationships among different historical snapshots to accomplish link prediction tasks [36]. GHT [32] involves two variants of Transformer and a relational continuous-time encoding function, aiming to mitigate the interference of facts irrelevant to the query and enhancing the capability of the model to learn long-term evolution. To effectively embed the rich information in the query-specific subgraph, SToKE [10] learns joint structural and temporal contextualized knowledge embeddings employing the pre-trained language model. Overview: While the Transformer framework has wide application in traditional KGs, its exploration in the field of TKGR remains relatively limited.

3 METHODOLOGY

3.1 Preliminary

Note that TKG $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{Q})$ denotes a directed multi-relation graph with timestamped edges between entities, where $\mathcal{E} \in \mathbb{R}^{|\mathcal{E}| \times d}$, $\mathcal{R} \in \mathbb{R}^{|\mathcal{R}| \times d}$, and $\mathcal{T} \in \mathbb{R}^{|\mathcal{T}| \times d}$ are the set of entities, relations and timestamps, respectively. $\mathcal{Q} = \{(s, p, o, \tau) \mid s, o \in \mathcal{E}, p \in \mathcal{R}, \tau \in \mathcal{T}\}$ is a set of quadruples in \mathcal{G} , where s, o, p and τ are a subject entity, an object entity, the relation between them, and a timestamp, respectively. For the link prediction task, we infer the missing object entity by converting $(s, p, ?, \tau)$. Similarly, we infer the missing subject entity by converting $(o, p^{-1}, ?, \tau)$, where we employ a reciprocal predicate p^{-1} to distinguish this case.

Evolutionary Chain of Events. The evolutionary chain of events (ECE) transforms structural and temporal contexts related to the query-specific neighborhood subgraph into a structured knowledge sequence. Specifically, for the query $(s, p, ?, \tau)$, ECE is conceptualized as a singly linked list $C = \{C_0, C_1, \dots, C_k\}$, comprising a chronologically arranged collection of events relevant to the subject s of the query. Notably, each C represents an event related to s , depicted by a triplet (e, p, τ) , where C_0 signifies the query and C_k denotes adjacent facts, k being the maximum adjacency count, e can serve as either the subject entity or the object entity. For example, for the query $(George\ Bush, visitTo, ?, 2002)$ as shown in Figure 1a, the ECE can be formulated as: $C = \{(George\ Bush, visitTo, 2002), (USA, presidentOf, 2001), (Barack\ Obama, succeeded, 2009)\}$. The formal expression is illustrated in Figure 1b. Hence, Our ECE can structurally represent both intra-quadruples and inter-quadruples. By encapsulating the dynamic evolution of entities and their inter-relations within the event chain, the ECE can capture the historical and current events related to the subject entity.

3.2 Overview of ECEformer

As an innovative end-to-end Transformer-based network, ECEformer learns the evolutionary chain of events (ECE) for TKGR. This work defines the subgraph composed of adjacent nodes that evolve over time along with the query entity as the evolutionary chain of events (ECE). Different from existing Transformer-based models, ECEformer crafts specific modules for handling intra-quadruples and inter-quadruples, respectively. As for intra-quadruples, ECE representation learning (ECER) based on a Transformer encoder is designed to explore the internal structure and semantic relationship within individual quadruples from the input ECE. On the other

hand, as for the inter-quadruples, mixed-context knowledge reasoning (MCKR) involves a novel information mixing layer, and the iterative transposition mechanism is designed to explore the contextual and temporal correlations among different quadruples. The architecture of ECEformer is shown in Figure 2, and more details will be illustrated in the following sections.

3.3 Learning Representation for Evolutionary Chain of Events

The first key to effectively learning ECE lies in the comprehensive extraction and analysis of each triplet corresponding to the sub-nodes on the ECE. Every constituent node of the ECE stores a historical event associated with the central entity, necessitating an investigation into its inherent structural and temporal characteristics. Inspired by the effectiveness of the Transformer architecture in learning inter-data dependencies, we introduce an ECER module utilizing a Transformer encoder that distills the embedding of each triplet. The details are shown in Figure 2c. Specifically, ECER consists of N core units, each consisting of alternating layers of multi-head attention (MHA), layer normalization (LN), and MLP-based feed-forward (FF). Among them, both MHA and FF are followed by a LN accompanied by the residual connection (RC). The FF contains two layers with the GELU non-linear activation function. This process can be formulated as:

$$y = l \circ r \circ f \circ l \circ r \circ h(x; W_h, W_l, W_f), \quad (1)$$

where h, l and f denote MHA function with weight W_h , LN function with weight W_l , and FF function with weight W_f , respectively. r denotes the RC function, which performs an operation that directly combines the input with the output of the intermediate layer. x denotes the input sequence. y denotes the output embedding. \circ denotes the composite function operator. Note that the MHA relies on the scaled dot-product attention mechanism to learn semantic interactions between the current token and all tokens within the sequence, thereby thoroughly exploring the structural and semantic information within individual events. Additionally, by directly inputting timestamps as tokens into ECER, the model can semantically extract temporal information within individual events.

Given a query $q(s, p, ?, \tau)$ and its ECE C_q , we treat each component within the branch of ECE as an individual token, sequentially unfolding the chain structure based on temporal order. This process yields a new input, as illustrated in Figure 2a. For each triplet on every branch of the ECE, we append a special token, i.e., $[CLS]$. Subsequently, we employ the concatenation of semantic embeddings E and positional embeddings P in the embedding space to represent each input token, as illustrated in Figure 2b. This process can be formulated as:

$$E_{inp}^i = E_i + P_j, \{i \in \mathcal{E} \cup \mathcal{R} \cup \mathcal{T}, j \in [0, 1, 2, 3]\}, \quad (2)$$

where E_{inp}^i denotes i -th token in the input sequence, E_i denotes the feature embedding of i -th token, and P_j denotes the position embedding. These embeddings are iteratively fed into the ECER module $\mathcal{F}_{ECER} : \mathbb{R}^{4 \times d} \rightarrow \mathbb{R}^d$. Eventually, the output consists of intermediate embeddings $E_{triple}^i, i = 0, 1, \dots, k$, which represent the structural and semantic information implied in the triples corresponding to each branch.

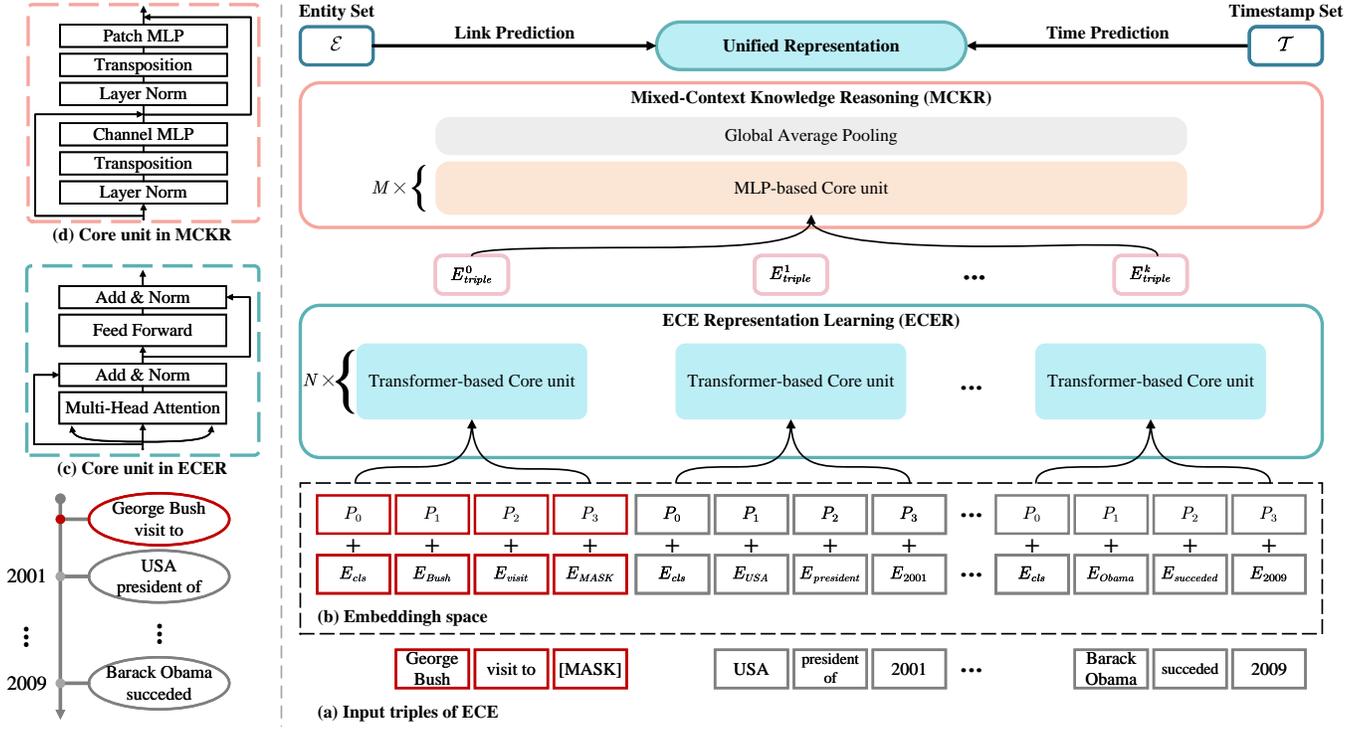


Figure 2: The architecture of the proposed ECEformer. Given an evolutionary chain of events, ECEformer obtains input sequences (detailed in (a)) from each branch of the chain and converts each token into a concatenation of semantic embeddings E and positional embeddings P (detailed in (b)). Subsequently, these branch embeddings E_{triple} are derived by the ECER, which is based on the Transformer encoder (detailed in (c)). The contextual information from different branch embeddings is then inductively processed by the MLP-based MCKR (detailed in (d)), culminating in a unified representation.

3.4 Mixed-Context Knowledge Reasoning

After encoding through the ECER, the embedded representations have effectively abstracted the relevant information of individual events. We introduce the query-specific contextual information to further enhance the accuracy of reasoning. Inspired by [34], we introduce the MLP-based MCKR module to learn the multi-dimensional information interaction of the inputs by leveraging two mixing strategies. The details are illustrated in Figure 2d.

Different from the single event input in ECER, the MCKR module processes the sequence of $1 + k$ embedding representations of ECE. This forms a two-dimensional real-valued input matrix $\mathcal{M} \in \mathbb{R}^{(1+k) \times d}$. The MCKR consists of M core units, each consisting of two key mixing layers: the channel MLP and the patch MLP. Additionally, each mixing unit is preceded by a layer normalization and a transposition layer. The channel MLP operates on the columns of \mathcal{M} , mapping each column from \mathbb{R}^{1+k} to \mathbb{R}^{1+k} . It aims to facilitate information interaction across different event embeddings within the same feature dimension. Conversely, the patch MLP operates on the rows of \mathcal{M} , mapping each row from \mathbb{R}^d to \mathbb{R}^d . It aims to facilitate information interaction within the same event embedding across different feature dimensions. This process can be formulated as:

$$\dot{\mathcal{M}}_{*,i} = \mathcal{M}_{*,i} + W_2 \sigma(W_1 \text{Norm}(\mathcal{M}_{*,i})), \text{ for } i = 1, 2, \dots, d \quad (3)$$

$$\ddot{\mathcal{M}}_{j,*} = \dot{\mathcal{M}}_{j,*} + W_4 \sigma(W_3 \text{Norm}(\dot{\mathcal{M}}_{j,*})), \text{ for } j = 1, 2, \dots, k+1 \quad (4)$$

where σ is the non-linear activation function (GELU). Each MLP block in MCKR contains two fully-connected layers, W_1 and W_2 are hidden weights in the channel MLP, W_3 and W_4 are hidden weights in the patch MLP.

Following the methodology outlined in [34], our MCKR employs the same parameter to translate every column (row) for channel MLP (patch MLP). This parameter sharing mechanism endows the network with positional invariance, enhancing its ability to proficiently process sequential data. Moreover, this approach also circumvents the substantial increase in model complexity that typically accompanies dimensional augmentation. Subsequently, we leverage a global average pooling layer to distill the mixed-context $\ddot{\mathcal{M}}$, which can be formulated as:

$$\mathcal{U}_{ECE} = \text{GlobalAvgPool}(\ddot{\mathcal{M}}), \quad (5)$$

where \mathcal{U}_{ECE} denotes the unified representation of ECE. Given a quadruple with either the subject or the object missing, the task of link prediction can be accomplished by computing the similarity between \mathcal{U}_{ECE} and all candidate entities. Link prediction can be mathematically formulated as:

$$p(e_{gt}|q) = \text{Softmax}(\text{Sim}(\mathcal{U}_{ECE}, \mathcal{E})), \quad (6)$$

Table 1: Statistics of the Experimental Datasets.

Dataset	$ \mathcal{E} $	$ \mathcal{R} $	$ \mathcal{T} $	Time	N_{train}	N_{valid}	N_{test}
GDEL T	500	20	366	15 mins	2,735,685	341,961	341,961
ICEWS05-15	10,488	251	4,017	24 hours	386,962	46,092	46,2775
ICEWS18	23,033	256	7,272	24 hours	373,018	45,995	49,545
ICEWS14	7,128	230	365	24 hours	63,685	13,823	13,222
YAGO11K	10,623	10	73	1 year	16,406	2,050	2,051
Wikidata12K	12,554	24	84	1 year	32,497	4,062	4,062

where Sim denotes the similarity function between \mathcal{U}_{ECE} and a candidate entity, q denotes the predicted result, obtained by selecting the $top - 1$ entity based on similarity scores. $p(e_{gt}|q)$ denotes the occurrence probability that q is the target entity e_{gt} .

3.5 Enhancement of Temporal Information

Considering that events in TKGs often coincide with specific timestamps, the unified representations learned from link prediction might be trivial solutions facilitated by contextual information. Such representations will introduce spurious correlations, because they only consider entities and relations while neglecting the temporal aspect of events. To address this issue, we propose a marked time prediction task to enhance the exploration of temporal information during the contextualization process. Specifically, we utilize a special token $[MASK]$ to replace the timestamp τ in the query-specific branch of ECE. In other words, we transform the $C_0(s, p, \tau)$ in ECE to $C_0(s, p, [MASK])$. This direct masking strategy introduces perturbations to the original timestamps, thereby prompting the model to learn contextual information. To promote the model’s capability of understanding temporal information, we train the model using masked inputs to recover the perturbed timestamp. This design can encourage the model to assimilate contextual information and deduce the occurrence time of the query-specific event. The methodology is analogous to link prediction, where we directly utilize the unified representation \mathcal{U}_{ECE} to predict the correct timestamp. The prediction probability $p(\tau_{gt}|q)$ can be formulated as:

$$p(\tau_{gt}|q) = \text{Softmax}(Sim(\mathcal{U}_{ECE}, \mathcal{T})), \quad (7)$$

where $p(\tau_{gt}|q)$ denotes the probability that predicted timestamp q corresponds to the target timestamp τ_{gt} .

3.6 Training and Optimization

For link prediction and time prediction tasks, we employ the cross-entropy function to calculate the loss during the training process. The loss function can be formulated as:

$$\mathcal{L} = - \sum_{(s,p,o,\tau) \in \mathcal{G}} \log(p(e_{gt}|q)) + \lambda \log(p(\tau_{gt}|q)), \quad (8)$$

where $(s, p, o, \tau) \in \mathcal{G}$ represents the historical events in the training set, λ weights the time prediction task.

Additionally, for diversifying our training data and reducing memory cost, we implement an ECE sampling strategy similar to the edge dropout regularization [6]. This approach samples only a portion of the neighborhood information of the query subject and filters out the ground truth target entity from the sampled neighborhood information.

4 EXPERIMENTS

In this section, the experimental settings are first introduced from four aspects, including datasets, evaluation metrics, compared baselines, and implementation details. Then, we comprehensively analyze the proposed ECEformer also from three aspects, i.e., **superiority**, **effectiveness**, and **sensitivity**.

4.1 Experimental Setup

4.1.1 Datasets and Evaluation Metrics. We evaluate the reasoning performance of our ECEformer on six TKG benchmark datasets, including GDEL T, ICEWS05-15, ICEWS18, ICEWS14, YAGO11K, and Wikidata12K. According to [25], we also split the datasets of GDEL T and ICEWS05-15/18/14 into train/valid/test by time-stamps. According to [8], we split the datasets of YAGO11K and Wikidata12K into train/valid/test by time intervals. The statistical information of datasets is shown in Table 1, in which *Time* represents time granularity. Two widely used evaluation metrics are adopted to quantify the performance, namely *MRR* and *Hits@k*. *MRR* represents the mean reciprocal rank of the inferred true entity among the queried candidates, and *Hits@k* measures the proportion of instances where the true entity appears among the top k ranked candidates.

4.1.2 Compared Baselines. Considering that different methods are validated on various datasets, we divided our experiments into two groups according to dataset size for a fair comparison. We compared with twenty state-of-the-art (SOTA) models. Specifically, the extrapolation baselines for GDEL T, ICEWS05-15, and ICEWS18 consist of RL-based models (TITer [33], TLogic [26], and DREAM [49]), GNN-based models (RE-GCN [24], TiRGN [21], HGLS [46], L^2 TKG [45], and RPC [25]), and Transformer-based models (GHT [32] and SToKE [10]). The interpolation baselines chosen for smaller datasets (ICEWS14, YAGO11K, and Wikidata12K) include TimePlex [15], TeLM [38], TeRo [40], RotateQVS [5], SANe [22], NeuSTIP [31], QDN [35], LCGE [28], TGeomE [41], and HyIE [47].

4.1.3 Implementation Details. All experiments are conducted on a single NVIDIA RTX A6000. The implementation¹ of our model is built upon the open-source framework LibKGE². The dimension of the input embedding is set to 320. The hidden widths of the feed-forward layer inside ECER and the channel (patch) MLP inside MCKR are set to 1024. The maximum neighborhood of ECE is set to 50. We configure the core unit numbers for ECER and MCKR to be $N = 3$ and $M = 6$, respectively. The activation function in our network is GELU.

We train our model via the Adamax [18] with an initial learning rate of 0.01 and an L2 weight decay rate of 0.01. We employed the warm-up training strategy, wherein the learning rate commenced from zero and increased linearly during the initial 10% of the training steps, followed by a linear decrease throughout the remaining training process. We train the model with a batch size of 512 for a maximum of 300 epochs, employing an early stopping mechanism based on MRR in the validation set. Detailed discussions on the effects of the weight coefficient λ , the number of core units N and M are in the subsection **Sensitivity Analysis**.

¹<https://github.com/seeyourmind/TKGELib>

²<https://github.com/uma-pi1/kge>

Table 2: Performance of link prediction on GDELТ, ICEWS05-15, and ICEWS18. The best results are marked in bold, the second results are marked by underlining.

Model	GDELТ				ICEWS05-15				ICEWS18			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TITer	18.19	11.52	19.20	31.00	47.60	38.29	52.74	64.86	29.98	22.05	33.46	44.83
TLogic	19.80	12.20	21.70	35.60	46.97	36.21	53.13	67.43	29.82	20.54	33.95	48.53
DREAM	28.10	19.30	31.10	44.70	56.80	47.30	65.10	78.60	<u>39.10</u>	<u>28.00</u>	<u>45.20</u>	62.70
RE-GCN	19.69	12.46	20.93	33.81	48.03	37.33	53.90	68.51	32.62	22.39	36.79	52.68
TiRGN	21.67	13.63	23.27	37.60	49.84	39.07	55.75	70.11	33.58	23.10	37.90	54.20
HGLS	19.04	11.79	-	33.23	46.21	35.32	-	67.12	29.32	19.21	-	49.83
L ² TKG	20.53	12.89	-	35.83	57.43	41.86	-	80.69	33.36	22.15	-	55.04
RPC	22.41	14.42	24.36	38.33	51.14	39.47	57.11	71.75	34.91	24.34	38.74	<u>55.89</u>
GHT	20.04	12.68	21.37	34.42	41.50	30.79	46.85	62.73	27.40	18.08	30.76	45.76
SToKE	<u>37.10</u>	<u>29.00</u>	<u>39.90</u>	<u>52.50</u>	<u>71.20</u>	<u>60.50</u>	<u>79.00</u>	88.50	-	-	-	-
ECEformer (Ours)	51.19	38.72	59.41	71.10	77.29	73.11	79.39	<u>85.17</u>	44.80	39.18	46.89	55.31

Table 3: Performance of link prediction task on ICEWS14, YAGO11K, and Wikidata12K. The best results are marked in bold, the second results are marked by underlining.

Model	ICEWS14				YAGO11K				Wikidata12K			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TimePlex	60.40	51.50	-	77.11	23.64	16.92	-	36.71	33.35	22.78	-	53.20
TeLM	62.50	54.50	67.30	77.40	19.10	12.90	19.40	32.10	33.20	23.10	36.00	54.20
TeRo	56.20	46.80	62.10	73.20	18.70	12.10	19.70	31.90	29.90	19.80	32.90	50.70
RotateQVS	59.10	50.70	64.20	75.40	18.90	12.40	19.90	32.30	-	-	-	-
SANe	63.80	55.80	68.80	78.20	25.00	18.00	<u>26.60</u>	40.10	<u>43.20</u>	<u>33.10</u>	48.30	64.00
NeuSTIP	-	-	-	-	<u>25.23</u>	<u>18.45</u>	-	37.76	34.78	24.38	-	53.75
QDN	64.30	56.70	68.80	78.40	19.80	13.10	20.10	32.80	-	-	-	-
LCGE	<u>66.70</u>	<u>58.80</u>	<u>71.40</u>	81.50	-	-	-	-	42.90	30.40	<u>49.50</u>	67.70
TGeomE	62.90	54.60	68.00	78.00	19.50	13.00	19.60	32.60	33.30	23.20	36.20	54.60
HyIE	63.10	56.30	68.70	78.60	19.10	12.50	20.10	32.60	30.10	19.70	32.80	50.60
ECEformer (Ours)	71.70	67.31	73.60	<u>80.30</u>	25.63	19.48	26.88	<u>37.84</u>	47.81	41.35	49.96	60.35

4.2 Performance Comparison for Superiority

We compare our model with 20 SOTA models ranging from 2020 to 2023 on six benchmark datasets. We report the result in Table 2 and Table 3, where the best performances are highlighted in bold, while the second-best are marked by underlining. A noteworthy observation is that our ECEformer significantly outperforms other baseline models across all datasets in terms of *MRR* and *Hits@k*. Specifically, ECEformer achieves performance improvements over the second-best models by 14.09%, 6.09%, 5.70%, 5.00%, 0.40%, and 4.61% on *MRR*, respectively. Additionally, it achieves gains of 9.72%, 12.61%, 11.18%, 8.51%, 1.03%, and 8.25% on *Hits@1*, respectively. This significant margin of improvement demonstrates the superior performance of our ECEformer.

In particular, our model exhibits notable performance in four metrics on GDELТ: *MRR*, *Hits@1*, *Hits@3*, and *Hits@10*, achieving respective scores of 51.19%, 38.72%, 59.41%, and 71.10%. Compared with RL-based methods, our ECEformer outperforms DREAM by 23.09% in *MRR*, by 9.42% in *Hits@1*, by 28.31% in *Hits@3*, and by

26.4% in *Hits@10*, respectively. Compared with GNN-based methods, our ECEformer outperforms RPC by 28.78% in *MRR*, by 24.30% in *Hits@1*, by 35.05% in *Hits@3*, and by 32.77% in *Hits@10*, respectively. Compared with Transformer-based methods, our ECEformer outperforms SToKE by 14.09% in *MRR*, by 9.72% in *Hits@1*, by 19.51% in *Hits@3*, and by 18.6% in *Hits@10*, respectively. This indicates that our model can effectively learn the unified representations from frequently changing historical events in datasets characterized by fine-grained temporal granularity. Based on the comparative results from the ICEWS (i.e., ICEWS05-15/18/14), ECEformer consistently surpasses all the baselines on the *MRR*, *Hits@1*, and *Hits@3* metrics. However, it lags behind the second-best model in *Hits@10*. This situation also occurs in the YAGO11K and Wikidata12K. This suggests that although our model demonstrates high precision in predictions, there is still room for improvement in its recall capability. Besides, compared to other datasets, our method shows a relatively slight performance improvement on YAGO11K. This is attributed to 1) the issue of missing timestamps in the YAGO11K

Table 4: Ablation study on different module combinations on ICEWS14 and Wikidata12K. TP represents the time prediction task. The best results are marked in bold.

Model	ICEWS14		Wikidata12K	
	MRR	Hits@10	MRR	Hits@10
ECER	70.60	80.40	43.49	57.08
ECEformer (w/o TP)	69.4	79.1	46.12	58.54
ECEformer	71.70	80.30	47.81	60.35

dataset, and 2) the NeuSTIP method additionally introduces Allen algebra [31] to refine temporal rules and designs a specialized time prediction evaluation function to capture temporal information. Nevertheless, the performance gain achieved by our method indicates that the proposed masked time prediction task can mitigate challenges arising from low data quality.

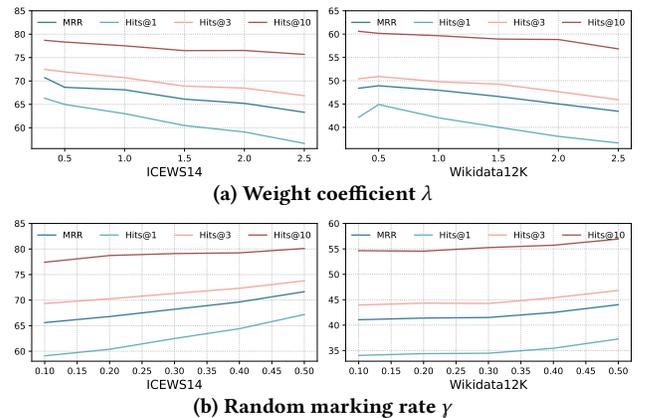
4.3 Ablation Study for Effectiveness

To analyze the contribution of ECEformer components, we validate them and their variants on the ICEWS14 and Wikidata12K: (1) using only ECER to embed individual query-specific quadruples (without contextual information); (2) combining ECER and MCKR without conducting the time prediction task; (3) combining ECER and MCKR while undertaking the time prediction task (i.e., the complete ECEformer). The detailed results are listed in Table 4.

From the results of ECER, we observe that by utilizing the Transformer architecture, our model can surpass most geoinformation-based models (such as TGeomE). This validates our viewpoint, which emphasizes sufficiently exploring the internal structure and semantic relationships within individual quadruples. Comparing the results of the ECER and ECEformer (w/o TP), we observe that incorporating contextual information significantly improves the model’s reasoning accuracy on Wikidata12K. However, this integration within ICEWS14 slightly diminishes performance. Specifically, ECEformer (w/o TP) outperforms ECER by 2.63% on Wikidata12K, but underperforms by 1.20% on ICEWS14 in terms of *MRR*. This phenomenon is attributed to the finer temporal granularity of ICEWS14 compared to Wikidata12K, where the frequent historical events in the context exacerbate spurious correlations. Without the intervention of the time prediction task, the model is more prone to erroneous estimations. Comparing the results of ECEformer (w/o TP) and ECEformer, we observe consistent performance enhancements on both datasets after executing the time prediction task. This is especially pronounced in ICEWS14, where the implementation of the time prediction task results in a notable augmentation of the *MRR*, surpassing the performance of the ECER. Specifically, ECEformer outperforms ECER by 1.10%, and it outperforms ECEformer (w/o TP) by 2.30% in terms of *MRR*. These findings confirm the effectiveness of our proposed temporal information enhancement strategy in prompting the model to leverage temporal information within its context, thereby effectively mitigating the impacts of spurious correlations.

Table 5: Sensitivity analysis of ECEformer with different number of unit layers. N and M represent the number of ECER units and MCKR units, respectively.

N	M	ICEWS14		Wikidata12K	
		MRR	Hits@10	MRR	Hits@10
1x		66.30	78.23	47.94	59.33
2x	1	67.36	78.77	48.08	60.00
3x		66.95	78.79	48.47	60.24
	1x	66.30	78.23	47.94	59.33
1	2x	69.01	78.53	48.31	60.19
	3x	70.30	79.08	48.19	60.09
1x		69.04	79.10	48.37	60.45
2x	2	69.34	79.03	48.32	60.51
3x		69.36	78.94	48.34	60.61
	1x	69.04	79.10	48.37	60.45
2	2x	70.18	78.80	48.22	60.39
	3x	70.10	78.19	47.67	59.21

**Figure 3: Sensitivity analysis of weight coefficients λ and random masking rate γ on ICEWS14 and Wikidata12K.**

4.4 Sensitivity Analysis

As mentioned in *Implementation Details*, we explore the performance fluctuations of ECEformer caused by the structural control parameters N and M , as well as the weight coefficient λ . For structural control parameters N and M , we set the base of one parameter to $\{1, 2\}$, while the other is adjusted in multiples, specifically to $1\times, 2\times, \text{ and } 3\times$. As shown in Table 5, the results highlight the maximum value in each group in bold. Our observations reveal that different structural settings do introduce perturbations to the model, with the disturbance being relatively minor at $N : M = 1 : 2$. Based on this finding, we configure the architecture of ECEformer as $N = 3$ and $M = 6$ across all datasets. For the weight coefficient λ , we record the model’s performance across a range of $\{0.5, 1.0, 1.5, 2.0, 2.5\}$, and plotted these results in Figure 3a. The performance curve trend suggests that excessively weighting the time prediction task can

detrimentally affect the link prediction task, thereby reducing overall performance. Therefore, setting the weight coefficient between 0.5 and 1.0 offers a balanced improvement in model performance.

In addition, to investigate the impact of the masking mechanism in the time prediction task on model performance, we report the results of each batch of data under varying masking rates, with details as well as illustrated in Figure 3b. Specifically, we employ a fixed masking rate γ during the training phase, randomly selecting $\gamma \times \text{batch_size}$ samples per batch to mask their timestamps. We observe that as the masking rate increases, the model performance correspondingly improves. Additionally, compared to Wikidata12K, the performance of ICEWS14 exhibits greater fluctuations with varying masking rates. This observation indicates that datasets with finer temporal granularity are more sensitive to variations in masking rate. Notably, our final model is trained using a fully masked approach.

5 CONCLUSION

In this paper, we address the challenge of completing missing factual elements in TKG reasoning by introducing the ECEformer, a novel end-to-end Transformer-based reasoning model. ECEformer primarily comprises two key components: the evolutionary chain of events representation learning (ECER) and the mixed-context knowledge reasoning (MCKR). The ECER, utilizing the Transformer encoder, thoroughly explores the structural and semantic information of historical events corresponding to each branch in the evolutionary chain of events (ECE). The ECE is composed of neighborhood subgraphs of entity nodes unfolded in chronological order. The MCKR based on MLP relies on the channel and patch mixing strategy to facilitate the learning of contextual information interactions within the ECE. Moreover, an additional time prediction task and a time masking mechanism are employed to force the model to assimilate temporal information from the context. Comparative experiments with state-of-the-art methods validate the superiority of our approach. Furthermore, additional ablation studies demonstrate the effectiveness of each module proposed in our ECEformer.

ACKNOWLEDGMENTS

This research was supported by National Science and Technology Major Project (2022ZD0119204), National Science Fund for Distinguished Young Scholars (62125601), National Natural Science Foundation of China (62076024, 62172035).

REFERENCES

- [1] Ralph Abboud, Ismail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. 2020. BoxE: A Box Embedding Model for Knowledge Base Completion. In *Advances in Neural Information Processing Systems*. 9649–9661.
- [2] Maristella Agosti, Stefano Marchesin, and Gianmaria Silvello. 2020. Learning unsupervised knowledge-enhanced representations to reduce the semantic gap in information retrieval. *ACM Transactions on Information Systems* 38, 4 (2020), 1–48.
- [3] Farah Atif, Ola El Khatib, and Djellel Difallah. 2023. BeamQA: Multi-hop Knowledge Graph Question Answering with Sequence-to-Sequence Prediction and Beam Search. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 781–790.
- [4] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems*. 2787–2795.
- [5] Kai Chen, Ye Wang, Yitong Li, and Aiping Li. 2022. RotateQVS: Representing Temporal Information as Rotations in Quaternion Vector Space for Temporal Knowledge Graph Completion. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 5843–5857.
- [6] Sanxing Chen, Xiaodong Liu, Jianfeng Gao, Jian Jiao, Ruofei Zhang, and Yangfeng Ji. 2021. HittER: Hierarchical Transformers for Knowledge Graph Embeddings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 10395–10407.
- [7] Xiang Chen, Ningyu Zhang, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, and Huajun Chen. 2022. Hybrid Transformer with Multi-Level Fusion for Multimodal Knowledge Graph Completion. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 904–915.
- [8] Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha P. Talukdar. 2018. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2001–2011.
- [9] Songgaojun Deng, Huzefa Rangwala, and Yue Ning. 2020. Dynamic knowledge graph based multi-event forecasting. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1585–1595.
- [10] Yifu Gao, Yongquan He, Zhigang Kan, Yi Han, Linbo Qiao, and Dongsheng Li. 2023. Learning Joint Structural and Temporal Contextualized Knowledge Embeddings for Temporal Knowledge Graph Completion. In *Findings of the Association for Computational Linguistics*. 417–430.
- [11] Rishab Goel, Seyed Mehran Kazemi, Marcus A. Brubaker, and Pascal Poupart. 2020. Diachronic Embedding for Temporal Knowledge Graph Completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3988–3995.
- [12] Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2020. DyERNIE: Dynamic Evolution of Riemannian Manifold Embeddings for Temporal Knowledge Graph Completion. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 7301–7316.
- [13] Zhiwei Hu, Victor Gutierrez-Basulto, Zhiliang Xiang, Ru Li, and Jeff Pan. 2022. Transformer-based Entity Typing in Knowledge Graphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 5988–6001.
- [14] Zhiwei Hu, Victor Gutierrez-Basulto, Zhiliang Xiang, Ru Li, and Jeff Z Pan. 2023. HyperFormer: Enhancing entity and relation interaction for hyper-relational knowledge graph completion. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. 803–812.
- [15] Prachi Jain, Sushant Rathi, Soumen Chakrabarti, et al. 2020. Temporal Knowledge Base Completion: New Algorithms and Evaluation Protocols. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 3733–3747.
- [16] Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2020. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 6669–6683.
- [17] Jaehun Jung, Jinhong Jung, and U Kang. 2021. Learning to walk across time for interpretable temporal knowledge graph completion. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 786–795.
- [18] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the International Conference on Learning Representations*.
- [19] Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor Decompositions for Temporal Knowledge Base Completion. In *Proceedings of the International Conference on Learning Representations*.
- [20] Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving Validity Time in Knowledge Graph. In *Companion of the International World Wide Web Conferences*. 1771–1776.
- [21] Yujia Li, Shiliang Sun, and Jing Zhao. 2022. Tirgn: time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence*. ijcai. org, 2152–2158.
- [22] Yancong Li, Xiaoming Zhang, Bo Zhang, and Haiying Ren. 2022. Each snapshot to each space: Space adaptation for temporal knowledge graph completion. In *Proceedings of the International Semantic Web Conference*. 248–266.
- [23] Zixuan Li, Xiaolong Jin, Saiping Guan, Wei Li, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2021. Search from History and Reason for Future: Two-stage Reasoning on Temporal Knowledge Graphs. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 4732–4743.
- [24] Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021. Temporal knowledge graph reasoning based on evolutionary representation learning. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 408–417.
- [25] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Si-hang Zhou, and Xinwang Liu. 2023. Learn from Relational Correlations and Periodic Events for Temporal Knowledge Graph Reasoning. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1559–1568.
- [26] Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*,

- Vol. 36. 4120–4127.
- [27] Johannes Messner, Ralph Abboud, and İsmail İlkan Ceylan. 2022. Temporal Knowledge Graph Completion Using Box Embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 7779–7787.
- [28] Guanglin Niu and Bo Li. 2023. Logic and commonsense-guided temporal knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 4569–4577.
- [29] Ali Sadeghian, Mohammadreza Armandpour, Anthony Colas, and Daisy Zhe Wang. 2021. ChronoR: Rotation Based Temporal Knowledge Graph Embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 6471–6479.
- [30] Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling Relational Data with Graph Convolutional Networks. In *Proceedings of the European Semantic Web Conference*, Vol. 10843. 593–607.
- [31] Ishaan Singh, Navdeep Kaur, Garima Gaur, et al. 2023. NeuSTIP: A Neuro-Symbolic Model for Link and Time Prediction in Temporal Knowledge Graphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 4497–4516.
- [32] Haohai Sun, Shangyi Geng, Jialun Zhong, Han Hu, and Kun He. 2022. Graph Hawkes Transformer for Extrapolated Reasoning on Temporal Knowledge Graphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 7481–7493.
- [33] Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. 2021. TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 8306–8319.
- [34] Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al. 2021. Mlp-mixer: An all-mlp architecture for vision. *Advances in neural information processing systems* 34 (2021), 24261–24272.
- [35] Jiapu Wang, Boyue Wang, Junbin Gao, Xiaoyan Li, Yongli Hu, and Baocai Yin. 2023. QDN: A Quadruplet Distributor Network for Temporal Knowledge Graph Completion. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–13.
- [36] Jiapu Wang, Boyue Wang, Meikang Qiu, Shirui Pan, Bo Xiong, Heng Liu, Linhao Luo, Tengfei Liu, Yongli Hu, Baocai Yin, et al. 2023. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects. *arXiv preprint arXiv:2308.02457* (2023).
- [37] Ruijie Wang, Zheng Li, Dachun Sun, Shengzhong Liu, Jinning Li, Bing Yin, and Tarek Abdelzaher. 2022. Learning to sample and aggregate: Few-shot reasoning over temporal knowledge graphs. *Advances in Neural Information Processing Systems* 35 (2022), 16863–16876.
- [38] Chengjin Xu, Yung-Yu Chen, Mojtaba Nayyeri, and Jens Lehmann. 2021. Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2569–2578.
- [39] Chenjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Yazdi, and Jens Lehmann. 2020. Temporal knowledge graph completion based on time series gaussian embedding. In *Proceedings of the International Semantic Web Conference*. 654–671.
- [40] Chengjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Shariat Yazdi, and Jens Lehmann. 2020. TeRo: A Time-aware Knowledge Graph Embedding via Temporal Rotation. In *Proceedings of the International Conference on Computational Linguistics*. 1583–1593.
- [41] Chengjin Xu, Mojtaba Nayyeri, Yung-Yu Chen, and Jens Lehmann. 2023. Geometric Algebra Based Embeddings for Static and Temporal Knowledge Graph Completion. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2023), 4838–4851.
- [42] Zezhong Xu, Peng Ye, Hui Chen, Meng Zhao, Huajun Chen, and Wen Zhang. 2022. Ruleformer: Context-aware Rule Mining over Knowledge Graph. In *Proceedings of the International Conference on Computational Linguistics*. 2551–2560.
- [43] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for knowledge graph completion. *arXiv preprint arXiv:1909.03193* (2019).
- [44] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1294–1303.
- [45] Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023. Learning latent relations for temporal knowledge graph reasoning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 12617–12631.
- [46] Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023. Learning Long-and Short-term Representations for Temporal Knowledge Graph Reasoning. In *Proceedings of the ACM Web Conference*. 2412–2422.
- [47] Sensen Zhang, Xun Liang, Hui Tang, and Zhenyu Guan. 2023. Hybrid Interaction Temporal Knowledge Graph Embedding Based on Householder Transformations. In *Proceedings of the ACM International Conference on Multimedia*. 8954–8962.
- [48] Wen Zhang, Yushan Zhu, Mingyang Chen, Yuxia Geng, Yufeng Huang, Yajing Xu, Wenting Song, and Huajun Chen. 2023. Structure Pretraining and Prompt Tuning for Knowledge Graph Transfer. In *Proceedings of the ACM Web Conference*. 2581–2590.
- [49] Shangfei Zheng, Hongzhi Yin, Tong Chen, Quoc Viet Hung Nguyen, Wei Chen, and Lei Zhao. 2023. DREAM: Adaptive Reinforcement Learning based on Attention Mechanism for Temporal Knowledge Graph Reasoning. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1578–1588.